

## **BAB III**

### **METODE PENELITIAN**

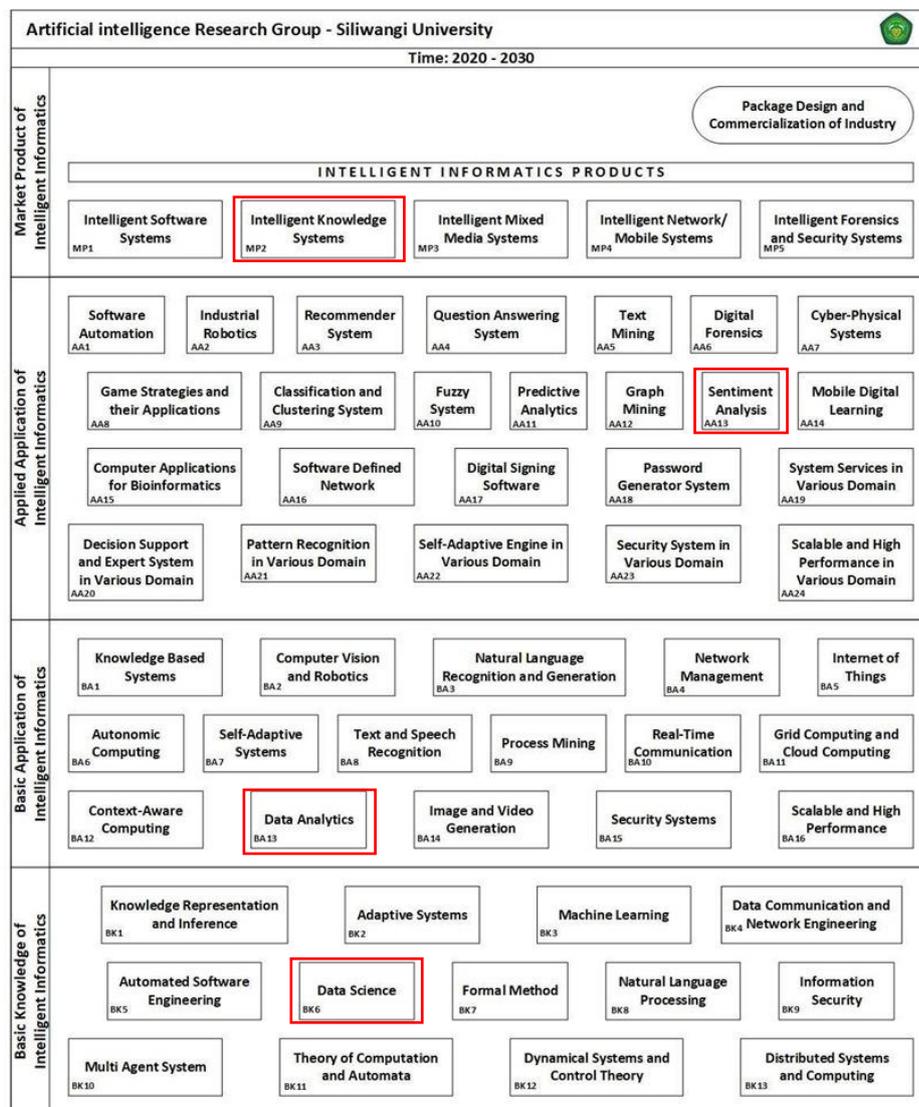
#### **3.1 Metodologi Penelitian**

Metode penelitian yang digunakan pada penelitian ini adalah penelitian kuantitatif yang berorientasi eksperimen. Menurut Hemawan (2005), penelitian kuantitatif merupakan suatu pendekatan penelitian yang bersifat objektif, mencakup pengumpulan dan analisis data kuantitatif serta menggunakan metode pengujian statistic (Laily, 2022). Metode penelitian kuantitatif dapat diartikan sebagai bagian dari serangkaian investigasi sistematis terhadap fenomena dengan mengumpulkan data untuk kemudian diukur dengan teknik statistik matematika atau komputasi (Fai, 2022).

Objek yang digunakan pada penelitian ini adalah komentar dari dua video YouTube yang relevan dengan Pilpres 2024 yang dipilih berdasarkan jumlah komentar yang banyak. Kedua video tersebut terdiri dari video yang diambil sebelum Pilpres 2024 dan setelah Pilpres 2024. Penelitian dilakukan dengan melakukan proses *preparation* data dan pemodelan dengan algoritma Naïve Bayes Classifier, yang kemudian akan diuji untuk memastikan akurasi dan efektivitasnya. Sentimen dari komentar pada kedua video akan diprediksi dan dianalisis untuk memahami perubahan opini publik sebelum dan sesudah Pilpre. Hasil analisis ini akan memberikan wawasan mendalam tentang dinamika sentimen masyarakat pada periode penting tersebut.

### 3.2 Peta Jalan (Road Map) Penelitian

Gambar 3.1 merupakan *Road Map* penelitian riset Kelompok Keahlian Sistem Inteligen (ISI) berkolaborasi dengan komunitas *AI Research Group Siliwangi University* yang dimana kajian sebelumnya dijadikan dasar pengetahuan untuk penyempurnaan bagi capaian akhir dari peta jalan penelitian secara keseluruhan.

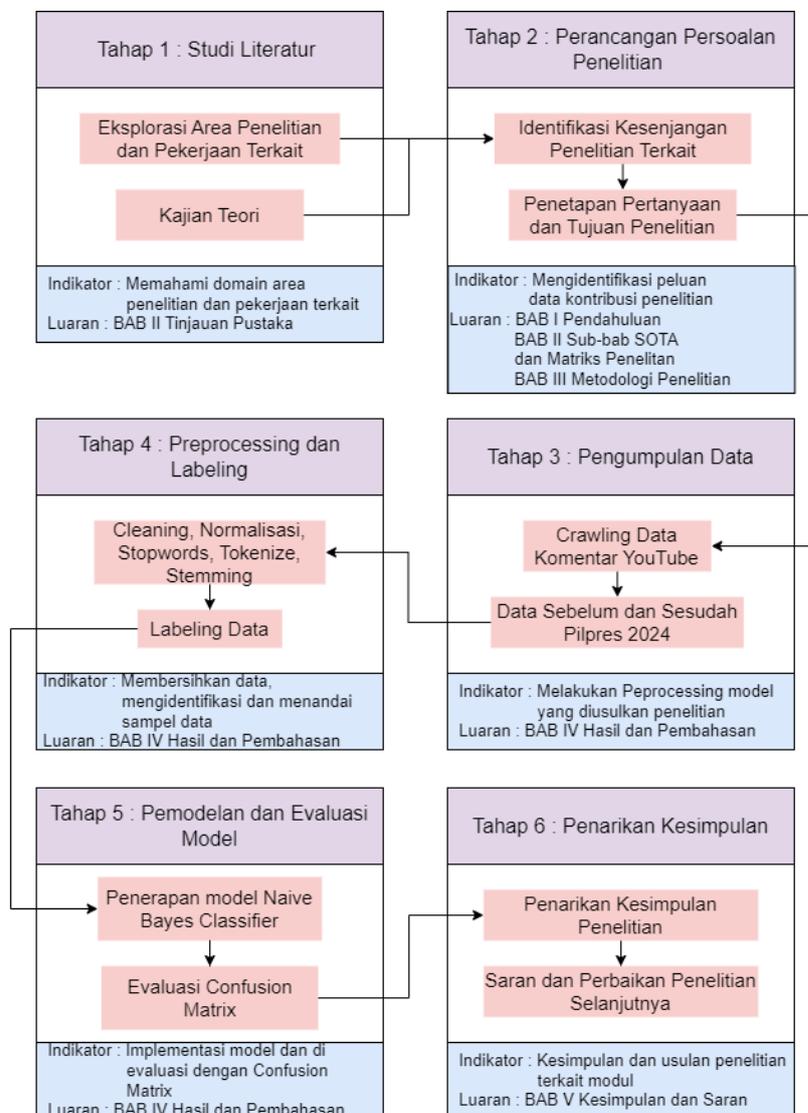


Gambar 3. 1 Road Map Penelitian Artificial Intelligence Siliwangi (AIS, 2021)

Gambar 3.1 menjelaskan tentang penelitian yang akan dilakukan yang ditandai dengan kolom berwarna merah yaitu terkait *data science*, *data analytics*, *sentimen analysis* untuk mencapai target berupa *intelligent knowledge system*.

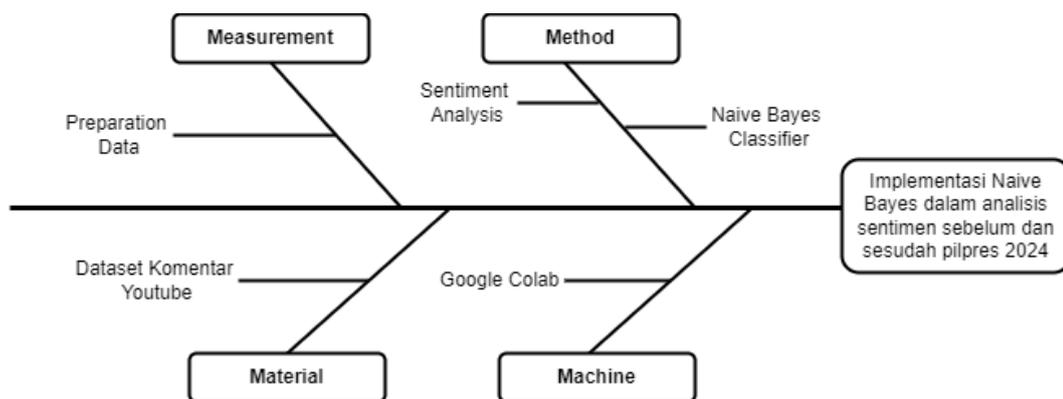
### 3.3 Tahapan Penelitian

Pada tahapan penelitian ini menjelaskan tahapan sistematis yang akan menjadi acuan pada kerangka penelitian. Gambar 3.2 merupakan alur penelitian yang terdiri dari 6 tahapan yaitu sebagai berikut.



Gambar 3. 2 Tahapan Penelitian

Tujuan penelitian ini juga direpresentasikan ke dalam Ishikawa Diagram atau biasa juga disebut *Fishbone* atau *cause-and-effect matrix* yang berisi penyebab-penyebab dari sebuah fenomena secara spesifik.



Gambar 3.3 Fishbone Diagram

Gambar 3.3 merupakan fishbone diagram dari penelitian ini mengkategorikan faktor-faktor tersebut ke dalam empat kelompok utama: Measurement, Method, Material, dan Machine.

a. Measurement (Pengukuran)

Preparation Data (Persiapan Data): Mengacu pada tahap pra-pemrosesan data, termasuk pengumpulan, pembersihan, dan normalisasi data komentar YouTube sebelum analisis sentimen dilakukan.

b. Method (Metode)

Sentiment Analysis (Analisis Sentimen): Proses menganalisis sentimen dari teks.

Naive Bayes Classifier: Algoritma yang digunakan untuk melakukan analisis sentimen pada komentar YouTube.

c. Material (Bahan)

Dataset Komentar YouTube: Sumber data yang digunakan dalam penelitian ini, yaitu komentar dari video YouTube yang relevan dengan Pilpres 2024.

d. Machine (Mesin)

Google Colab: Platform komputasi yang digunakan untuk menjalankan analisis dan pemodelan data, menyediakan lingkungan pemrograman berbasis cloud.

Diagram ini mengilustrasikan bagaimana masing-masing faktor berkontribusi terhadap keberhasilan implementasi Naive Bayes Classifier dalam analisis sentimen dengan tujuan utama memahami dinamika sentimen publik sebelum dan sesudah Pilpres 2024.

### **3.3.1 Studi Literatur**

Studi literatur merupakan proses sistematis mengumpulkan, mengevaluasi, dan meringkas literatur atau penelitian yang relevan dengan topik penelitian yang akan dibuat. Tujuan utama dari studi literatur yaitu untuk memahami keadaan terkini dari pengetahuan di bidang tertentu, mengidentifikasi celah pengetahuan, dan merumuskan kerangka konseptual untuk penelitian selanjutnya. Melakukan pencarian literatur menggunakan basis data akademis, perpustakaan digital, jurnal ilmiah, dan sumber-sumber lainnya yang relevan dengan topik penelitian.

### **3.3.2 Perancangan Persoalan Penelitian**

Perancangan persoalan penelitian merupakan tahapan yang dilakukan berdasarkan analisis penelitian sebelumnya. Masalah penelitian diidentifikasi dengan mengevaluasi kesenjangan, kekurangan, dan rekomendasi dari penelitian-penelitian terdahulu, sehingga perbaikan dapat dilakukan. Selanjutnya desain penelitian dibuat dengan menetapkan tujuan yang jelas melalui formulasi pertanyaan penelitian, batasan penelitian, dan manfaat penelitian.

### **3.3.3 Pengumpulan data**

Pengumpulan data untuk analisis sentimen melibatkan pengumpulan sejumlah besar data teks, seperti ulasan produk, komentar media sosial, atau entri blog, yang nantinya akan dianalisis untuk menentukan sentimen atau pendapat yang terkandung dalam teks tersebut. Crawling data (Perayapan data) merupakan suatu proses yang menggali jauh ke dalam world wide wet, atau target tertentu untuk mengambil data. Crawling data juga didefinisikan sebagai proses otomatis untuk mengumpulkan dan mengindeks data dari berbagai sumber. Aktivitas ini biasanya dilakukan oleh bot atau software yang disebut dengan crawler. Data yang diambil dari hasil crawling ini pada umumnya akan dianalisis, lalu dijadikan bahan pengembangan sistem, atau bahkan dijadikan data penelitian tertentu (Mahmud, 2023).

Pengumpulan data pada penelitian ini dilakukan dengan mengumpulkan komentar dari satu video Youtube yang relevan dengan Pemilihan Presiden Republik Indonesia 2024. Pengambilan data komentar dilakukan pada dua waktu yang berbeda. Data komentar YouTube sebelum Pilpres dikumpulkan setelah

proses debat selesai karena pada waktu tersebut, diskusi publik tentang calon presiden biasanya mencapai puncaknya. Hal ini mencerminkan opini masyarakat yang masih terfokus pada program, visi, misi, dan performa kandidat selama debat. Oleh karena itu, pengumpulan data pada momen ini dianggap paling representatif untuk menangkap sentimen publik sebelum pemilihan. Sementara itu, data komentar setelah Pilpres diambil setelah proses perhitungan suara selesai. Hal ini dilakukan karena pada waktu tersebut, hasil pemilihan sudah diketahui, dan sentimen publik mulai mencerminkan reaksi mereka terhadap hasil tersebut, baik berupa kepuasan, optimisme, maupun kekecewaan. Pengumpulan data pada momen ini memungkinkan analisis yang lebih akurat terhadap perubahan sentimen publik sebelum dan setelah Pilpres. Pendekatan ini dilakukan untuk memastikan bahwa perbandingan antara sentimen sebelum dan sesudah Pilpres didasarkan pada momen-momen yang signifikan dan relevan, sehingga dapat memberikan gambaran yang jelas mengenai perubahan opini publik selama periode tersebut. Berikut adalah satu video yang komentarnya dijadikan sebagai sumber data pada penelitian yang akan dilakukan:

Tabel 3. 1 Pengumpulan Data Komentar Video YouTube

<b>Judul Video</b>	<b>Keterangan</b>	<b>Link</b>
[Full] Panas! Debat Ketiga Pemilu 2024, Para Capres Saling Serang	7 Januari 2024 (Sebelum Pilpres 2024)	<a href="https://www.youtube.com/watch?v=XUE2j8dujPM">https://www.youtube.com/watch?v=XUE2j8dujPM</a>

### 3.3.4 *Preprocessing dan Labeling*

*Preprocessing* merupakan tahapan proses untuk membersihkan data dari komentar youtube yang tidak diperlukan serta kata-kata yang tidak memiliki makna. Proses ini dilakukan sesuai dengan isi data dari proses crawling data. Adapun tahapannya yaitu *cleaning, case folding, tokenizing, stemming dan stopword removal* (Ningtyas et al., 2023)

*Labeling* atau pelabelan data merupakan proses mengidentifikasi dan menandai sampel data yang biasa digunakan dalam konteks pelatihan model machine learning. Prosesnya bisa manual, tetapi biasanya dilakukan atau dibantu dengan perangkat lunak. Pelabelan data juga membantu model machine learning membuat prediksi yang akurat dan berguna dalam proses seperti visi computer, pemrosesan Bahasa alami dan pengenalan ucapan (Gillis, 2023).

### 3.3.5 *Pemodelan dan Evaluasi Model*

Pada tahapan ini dilakukan dengan pembagian dataset yang telah diproses menjadi data pelatihan dan data pengujian. Model Naïve Bayes dilatih menggunakan data pelatihan, di mana model mempelajari probabilitas kata-kata muncul dalam setiap kategori sentimen. Setelah pelatihan, model diterapkan pada data pengujian untuk memprediksi kategori sentimen dari teks yang belum dilihat sebelumnya.

Penggunaan teknik SMOTE (Synthetic Minority Oversampling Technique) diterapkan pada data pelatihan untuk mengatasi masalah ketidakseimbangan kelas yang dapat memengaruhi performa model. Dengan

SMOTE, sampel sintetis baru dibuat untuk kelas minoritas, sehingga jumlah data antara kelas mayoritas dan minoritas menjadi lebih seimbang. Hal ini membantu model Naïve Bayes mempelajari distribusi data secara lebih efektif dan mencegah bias terhadap kelas mayoritas.

Evaluasi model merupakan tahap kritis dalam proses pengembangan model machine learning. Hal ini memberikan gambaran seberapa baik model dapat memprediksi data yang belum pernah dilihat sebelumnya. Untuk model klasifikasi seperti Naive Bayes Classifier, evaluasi melibatkan perbandingan antara prediksi model dengan label aktual pada data uji. Beberapa metrik evaluasi umum yang digunakan yaitu confusion matrix, accuracy, precision, recall, dan F1-score.

### **3.3.6 Penarikan Kesimpulan**

Pada tahapan terakhir ini dilakukan dengan visualisasi hasil dari analisis sentimen atau klasifikasi menggunakan Naive Bayes Classifier bisa memberikan gambaran yang lebih nyata tentang performa model. Pemilihan jenis visualisasi tergantung pada tujuan spesifik dan sifat data yang di analisis. Visualisasi dapat membantu memudahkan interpretasi hasil, memberikan wawasan lebih dalam, dan mendukung pengambilan keputusan.